



Projeto e Implementação de Plataforma de Dados com Snowflake

Modern Data Stack
com Snowflake, Airbyte e dbt

Projeto e Implementação de Plataforma de Dados com Snowflake

O Modern Data Stack é um ecossistema de ferramentas modernas projetadas para lidar com os desafios de processamento, armazenamento e análise de dados de forma escalável, ágil e eficiente. Ele utiliza tecnologias baseadas na nuvem e práticas modernas de engenharia de dados. Um exemplo robusto desse stack envolve Snowflake, Airbyte e dbt, que desempenham papéis complementares, como veremos a seguir.

Snowflake é uma plataforma de armazenamento e processamento de dados baseada na nuvem. Ele combina as funcionalidades de um Data Warehouse, Data Lake e uma plataforma de compartilhamento de dados. Suas principais características incluem:

- Escalabilidade: Capacidade de ajustar os recursos de forma elástica, conforme necessário.
- Arquitetura baseada em nuvem: Totalmente desacoplado entre armazenamento e computação.
- Compatibilidade com SQL: Facilita a adoção e permite análises rápidas.
- Gestão simplificada: Reduz a necessidade de manutenção pesada, como tuning de hardware ou software.

Airbyte é uma plataforma de integração de dados de código aberto, projetada para extrair dados de várias fontes (como APIs, bancos de dados e ferramentas SaaS) e carregá-los em destinos como Snowflake. Ele é particularmente útil para a etapa EL (Extract-Load) do processo ETL/ELT. Destaques incluem:

- Conectores prontos para uso: Integra-se facilmente a centenas de fontes e destinos.
- Personalização: Permite criar conectores específicos para necessidades únicas.
- Automação: Pode agendar e monitorar pipelines de dados.
- Flexibilidade: Suporta o formato ELT, que é mais adequado para o Modern Data Stack.

O dbt é uma ferramenta voltada para a transformação de dados (o "T" em ELT), que opera diretamente dentro de um Data Warehouse, como o Snowflake. Ele é baseado em SQL e permite que engenheiros e analistas construam pipelines de dados limpos e organizados. Principais funcionalidades:

- Modelagem de dados declarativa: Facilita o uso de SQL para definir transformações.
- Versionamento e colaboração: Permite controle de versões de código e integração com ferramentas como Git.
- Teste de dados: Possui recursos integrados para validação de qualidade de dados.
- Automação de pipelines: Organiza e executa transformações em sequência.

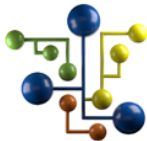
Projeto e Implementação de Plataforma de Dados com Snowflake

Como essas ferramentas se integram no Modern Data Stack

- Extração e carregamento com Airbyte: Os dados brutos são extraídos de diferentes fontes e carregados diretamente no Snowflake.
- Armazenamento e processamento no Snowflake: O Snowflake serve como o repositório central para os dados brutos, enquanto também fornece capacidade de processamento para consultas e análises.
- Transformação e modelagem com dbt: Os dados brutos carregados no Snowflake são transformados, limpos e modelados utilizando os scripts do dbt. Isso prepara os dados para análises avançadas.

Essa combinação oferece um fluxo de trabalho escalável e otimizado para atender às demandas modernas de dados, permitindo insights rápidos e maior flexibilidade operacional.

É exatamente isso que iremos construir na prática agora neste capítulo!



Equipe DSA

Muito Obrigado!
Continue Trilhando Uma Excelente Jornada de Aprendizagem.